MORPACE Post-Processing Data Checks

Data processing prepares SAS data files according to specifications provided by the research department (SAS is a data analysis system). The research department uses SAS to run post-processing data checks, ensuring that the information provided to MDOT is complete and accurate. The following pages itemize the MDOT MI Travel Counts data checks that are performed.

Household File

1. Every variable should have an answer. There should be no blanks.
2. Each case should have a record type of "1" for "Household Record".
3. Check that there are no duplicate QNOs or phone numbers.
4. Check that the area codes are valid Michigan area codes.
5. Check that the month/day combinations are correct and that they match the day of the week variable.
6. Perform a quick visual check for city misspellings.
7. Check that all home addresses are located in Michigan.
8. Check that all home zip codes are between 48000 and 49999.
9. Check that all counties listed are included in the master list of the 83 Michigan counties.
10. Check that home longitude is always a negative value, typically between 80 and 90.
11. Check that home latitude is always a positive value, typically between 40 and 50.
12. Check that longitude and latitude is not rounded. Data should be to six decimal places.
13. Check that the number of workers in the household does not exceed the number of people in the household.
14. Check that the number of subsidized vehicles does not exceed the number of vehicles available to the household.

- In the raw data file from the recruit, the region assignment is checked for each household.

Person File

1. Each case should have a record type of "2" for "Person Record".
2. QNO and phone number will have duplicates. Frequency of QNO and phone number is the number of persons in the household.
3. Person number will have duplicates.
4. Check person number against the number of people in the household from the household file.
5. Person number 1 frequency should be the total number of households.
6. Check that all cases have a value for age range.
7. Only respondents with AGERNG=98 OR 99 should have an answer for the AGE18 variable.
8. Check that all cases have a value for relationship.
9. Check that the contact person (code 0) is person number 1.
10. Spot check ages of a few relationship categories (spouses are probably not under 18 years of age).
11. Only respondents that said "other" for relationship should have an answer for other relationship to contact person.
12. Check that no cases are missing for licensed driver. Check that respondents 16 years of age or older are not code 3 for licensed driver. Check that respondents under 16 years of age are code 3 (not applicable).
13. Check that no cases are missing for transit pass. The majority of cases should be no.
14. Check that only respondents with a transit pass have one or more answers for type of transit pass.
15. Check that only respondents who indicated "other" transit pass have an answer for other type of transit pass.
16. Check that no cases are missing for education level. Check that respondents 18 years of age or older are not code 0 for education level. Check that respondents under 18 years of age are code 0 (not applicable).
17. Check that no cases are missing school type.

18. Check that no cases are missing for school name through school zone, if respondent is currently a student.
19. Check that almost all states for school address are Michigan and zip codes are between 48000 and 49999, unless student is above K-12 level.
20. Check that school longitude is always a negative value, typically between 80 and 90.
21. Check that school latitude is always a positive value, typically between 40 and 50.
22. Check that longitude and latitude is not rounded. Data should be to six decimal places.
23. Check that no cases are missing for working status. Check that respondents 16 years of age or older are not code 5 for working status. Check that respondents under 16 years of age are code 5 (not applicable).
24. Check that respondents that are not working are asked the not working status question.
25. Check that worker questions are only asked if working status is code 1 or code 2.
26. Check that only respondents who indicated "other" industry have an answer for other industry.
27. Check that only respondents who have a fixed workplace have answers for work address through work zone.
28. Check that almost all states for work address are Michigan and zip codes are between 48000 and 49999.
29. Check that work longitude is always a negative value, typically between 80 and 90.
30. Check that work latitude is always a positive value, typically between 40 and 50.
31. Check that longitude and latitude is not rounded. Data should be to six decimal places.
32. Check that secondary job questions are only asked if respondent has more than one job.
33. Check that no cases are missing proxy status.
34. Check that an infant or child is not a "respondent" interview. Interviewers are not allowed to talk directly with someone under 14 years of age. Interviewers are only allowed to talk directly to 14 and 15 year olds with parental approval.
35. Check that all proxy cases indicate which household member provided the proxy information.
36. Check that no proxy cases indicate that the person number providing the proxy is the same as the person number of the respondent.
37. Check that no cases are missing the diary completed variable.
38. Check that respondents who completed the diary are not code 3 (not applicable) for using the completed diary. Other respondents should be code 3 for using the completed diary.
39. Check that no cases are missing for long distance trips taken.

Person File Benchmarks

a. Distribution of gender should be approximately 50% male and 50% female. May have slightly more females than males.
b. More cases should be no than yes for primary job including evenings.
c. Most cases should be no for primary job including overnights.
d. A large percent of primary job weekly hours should be 40 hours.
e. Primary job flexibility should mostly be no or some flexibility.
f. Primary job compressed week should mostly be no.
g. Secondary job may be more flexible than primary job.
h. Secondary job weekly hours should be much lower than hours for primary job.

Trip File

1. Each case should have a record type of "3" for "Trip Record".
2. QNOs with a frequency of 1 are households that did not take any trips. There should be very few no trip households.
3. TRIPNUM=0 frequency is the number of no trip people. Again, there should be very few no trip people.
4. Cases that did not take any trips will only have origin information – where they started and ended the travel period.
5. Check that origin and destination longitude is always a negative value, often between 80 and 90.
6. Check that origin and destination latitude is always a positive value, often between 40 and 50.

7. Check that longitude and latitude is not rounded. Data should be to six decimal places.
8. Check that only respondents who indicated "other" for type of origin or destination have an answer for other type of location.
9. Check that the number of TRAV=2 is equal to the number of people in the study.
10. Check that all cases of TRAV=2 and TRIPNUM=0 provided a reason for no travel.
11. If the trip number is equal to 1, time of departure and type of transportation used should be answered, unless respondent began travel period traveling. If the trip number is greater than 1, time of departure and type of transportation used should be answered.
12. Check that only cases with "other" for type of transportation have an answer for other type of transportation used.
13. Check that the bus provider used is not missing if the trip involved dial-a-ride or a public bus as one of the transportation modes.
14. Check that only cases with "other" for bus provider have an answer for other bus provider used.
15. Check that there are no cases missing "pay for trip" if transportation types 6 (taxi/shuttle), 7 (dial-a-ride), 8 (train), or 9 (public bus) were used as a transportation mode.
16. Check that if the respondent indicated they paid for the trip that a valid amount has been recorded for amount paid for trip.
17. Check that if a trip involved a car, van, truck, or motorcycle that the driver/passenger variable is not missing. Note that children that are too young to drive (under 14 years of age) are not asked the question, but are postcoded as passengers.
18. Check that if a trip involved a car, van, truck, or motorcycle that the number of additional people in the vehicle was asked.
19. If the respondent did take a trip with other people in a car, van, truck, or motorcycle, check that the number of household members in the vehicle was asked, unless the respondent lives alone. If the respondent is a one-member household, the variable is postcoded with "none".
20. Check that the number of household members in the vehicle is not greater than the number of people in the vehicle.
21. Check that the respondent is not listed as a household member in the vehicle.
22. Check that if a trip involved a car, van, truck, or motorcycle that the respondent was asked if a household vehicle was used for the trip, unless the household does not have any available vehicles. If the respondent is a zero-vehicle household, the variable is postcoded with "no".
23. Check that if a trip involved a car, van, truck, or motorcycle that the respondent was asked if they paid for parking.
24. Check that if the respondent indicated they paid for parking that a valid amount has been recorded for amount paid for parking and that a parking rate has been identified. If the parking rate is "other", the answer should be recorded in the other parking rate variable.
25. Check that all cases (except those respondents that did not travel) are not missing arrival time or destination information.
26. Check that if the respondent was at home that only activity codes 1 or 2 are used.
27. Check that if the respondent was not a home that activity codes are not code 1 or 2.
28. Check that the departure and arrival times are in military time.

Long Distance File

1. Each case should have a record type of "4" for "Long Distance Trip Record".
2. The frequency of LDTRIP=1 should be the number of LDTRIPS=2 in the person file.
3. Check that only cases with "other" for type of transportation used to reach location have an answer for other type of transportation used.
4. Check that the bus provider used to reach the location is not missing if a public bus was the mode of transportation used to reach the location.
5. Check that only cases with "other" for bus provider have an answer for other bus provider used to reach the location.
6. Check that only cases with "other" for type of transportation used at the location have an answer for other type of transportation used at the location.
7. Check that the bus provider used at the location is not missing if a public bus was used at the location.

8. Check that only cases with "other" for bus provider used have an answer for other bus provider used at the location.
9. Check that the number of times the trip was taken in the last 3 months is not greater than the number of times the trip has been taken in the last 12 months.
10. Check that the number of times the trip has been taken in the last 12 months is asked, unless the respondent indicated don't know or refused to the number of times the trip was taken in the last 3 months.

Any discrepancies or missing values are further investigated by the research department, in conjunction with data processing.

For the pilot, the following issues were faced:

SNAME (Person File)

SNAME was originally omitted from the data structure. This was an oversight by the research department. The variable was added and the data structure was updated accordingly prior to data delivery.

W1NAME and W1TYPE (Person File)

Originally, there were 11 cases missing these variables. These were the 11 cases that reported working only at home or having no fixed workplace (W1LOC=2 OR 3). They were asked W1NAME and W1TYPE, but the data processing department had not included the information in the original data file. The problem was identified and corrected before the pilot data was delivered.

NO TRAVEL PERSONS (Trip File)
For the 25 persons (TRIPNUM=0) that did not travel during the 48-hour travel period, the ORIGIN to ACT4 variables were originally blank in the data file. These variables were updated in the data file so that MDOT can see at what location the respondent spent the travel period.

VHNUM (Trip File)
There were 25 cases missing originally. These were cases where the household is a one-member household. The cases were correctly post-coded as "0" prior to data delivery.

DTIME and ATIME (Trip File)
The values for these variables are to be delivered in military time. The data was originally delivered with a combination of am/pm and military times. The research department did not check for or realize the mistake prior to data delivery, which was an oversight. A check has been added to the post-processing data checks to ensure that this mistake will not happen in the future.